SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 78-0276	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATING DEPENDENT LIFE LENGTHS, WITH APPLICATIONS TO THE Theory COMPETING RISKS		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER FSU-~~No~~-M438
7. AUTHOR(s) N. Langberg, F. Proschan and A. J. Quinzi		8. CONTRACT OR GRANT NUMBER(s) AFOSR 74-2581
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Florida State University Department of Statistics Tallahassee, FL 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332		12. REPORT DATE November 1977
		13. NUMBER OF PAGES 30 35 p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Competing risks, life distribution, simultaneous failure, reliability, Kaplan-Meier estimator, consistent estimator, survival probability, dependent life lengths, censored data, dependent components, estimating risks.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In the classical theory of competing risks (as well as in many reliability models and incomplete data problems) it is assumed that ~~(a)~~ the risks (i.e., the random variables of interest) are independent and that ~~(b)~~ death does not result from simultaneous causes. Employing ~~our~~ the probabilistic solution to a related problem in probability modelling, ~~we obtain~~ strongly consistent estimators are obtained for the unobservable marginal distributions of interest. These estimators are analogous to those of Kaplan and Meier [J. Amer. Statist. Assoc. (1958) 63] and are appropriate when the assumptions of independence and no

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract

simultaneous causes of death [(a) and (b), above] fail to hold. ~~We show how our~~ These methods can be used to unify and simplify the nonparametric approach toward estimation in the competing risks model. As a consequence we obtain an elementary proof of the strong consistency of the Kaplan-Meier estimator. Our results extend and simplify the work of Peterson [J. Amer. Statist. Assoc. (1977) 72] and Desu and Narula [The Theory and Applications of Reliability, I, ed. by I. Shimi and C. P. Tsokos (1977)].

AFOSR-TR- 78-0276

18 AFOSR

19 TR-78-Ø276

6

ESTIMATING DEPENDENT LIFE LENGTHS, WITH APPLICATIONS TO THE THEORY OF COMPETING RISKS.

9 Interim rept., by

10 N./Langberg[redacted], F./Proschan[redacted] [redacted] A. J./Quinzi[redacted]

11 Nov 77

Statistics Report M438,
AFOSR Technical Report No. 78

14 FSU-STATISTICS-M438, TR-78-AFOSR

November, 1977
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

15 ✓AFOSR-74-2581

12 35p.

16 23Ø4

17 A5

[1] The Florida State University

[2] Temple University

[3] Research supported by the Air Force Office of Scientific Research under AFOSR Grant 74-2581D.

Approved for public release;
distribution unlimited.

400 277

Estimating Dependent Life Lengths, with Applications to the Theory of Competing Risks

by

N. Langberg[1], F. Proschan[1,3], and A. J. Quinzi[2]

ABSTRACT

In the classical theory of competing risks (as well as in many reliability models and incomplete data problems) it is assumed that (a) the risks (i.e., the random variables of interest) are independent and that (b) death does not result from simultaneous causes. Employing our probabilistic solution to a related problem in probability modelling, we obtain strongly consistent estimators for the unobservable marginal distributions of interest. These estimators are analogous to those of Kaplan and Meier [J. Amer. Statist. Assoc. (1958) 63] and are appropriate when the assumptions of independence and no simultaneous causes of death [(a) and (b), above] fail to hold. We show how our methods can be used to unify and simplify the nonparametric approach toward estimation in the competing risks model. As a consequence we obtain an elementary proof of the strong consistency of the Kaplan-Meier estimator. Our results extend and simplify the work of Peterson [J. Amer. Statist. Assoc. (1977) 72] and Desu and Narula [The Theory and Applications of Reliability, I, ed. by I. Shimi and C. P. Tsokos (1977)].

---

[1]The Florida State University

[2]Temple University

[3]Research supported by the Air Force Office of Scientific Research under AFOSR Grant 74-2581D.

1. Introduction and summary.

Langberg, Proschan, and Quinzi (1978) [hereafter referred to as LPQ (1978)] show that under certain mild conditions it is possible to establish a particular equivalence between an arbitrary system of dependent components and a system of independent components. The two systems are equivalent in the sense that they possess (1) the same distribution for the time to system failure and (2) the same probabilities of occurrence of each failure pattern. In the case of a series system, a particular failure pattern occurs when the failure of a particular set of components coincides with the failure of the system. The LPQ (1978) result extends results by Miller (1977) and Tsiatis (1975). In Section 2 we state the LPQ (1978) result. Although formulated in reliability terms, the result applies to any model where observations consist of (1) the time at which a particular event occurs and (2) the identity of the cause or combination of causes (among a finite number) which results in the occurrence of the event. In particular, in population mortality studies, the data on each subject includes (1) the age at death and (2) the cause(s) of death. In this case, an individual dies due to the occurrence of one or more of a finite number of possible causes of death, which are sometimes viewed as "competing" for the individual's life. Moreover, one or more of the "causes" might be identified with withdrawal of an individual from observation [referred to as a "loss"], resulting in censored or truncated data. Thus, the LPQ (1978) result, as well as the methods developed in this paper, apply in at least three contexts of interest--(1) engineering or reliability models, (2) competing risks models, and (3) various models involving incomplete data.

In the classical theory of competing risks it is assumed that the risks act independently and that death does not result from simultaneous causes. In Section 3 we describe the competing risks model. We examine the classical assumptions and introduce the Kaplan-Meier (1958) (K-M) estimator. Employing the LPQ (1978) result, we obtain in Section 4 strongly consistent estimators for the unobservable marginal distributions of interest in the competing risks model. These estimators are analogous to those of Kaplan and Meier (1958) and are appropriate when the assumptions of independence and no simultaneous causes of death fail to hold. We show how our methods can be used to unify and simplify the nonparametric approach toward estimation in the competing risks model. As a consequence we obtain an elementary proof of the strong consistency of the Kaplan-Meier estimator.

Our results extend and simplify the work of Desu and Narula (1977) and Peterson (1975, 1977). Section 5 consists of proofs.

2. <u>From a dependent model to an independent one</u>.

In this section we state, for future reference, the equivalence result of LPQ (1978).

To state our model precisely, it is convenient to use the language of reliability theory (system, component, etc.) although the result is applicable in a variety of other contexts, especially in the context of competing risks (see Section 3). A <u>life length</u> T is a nonnegative random variable such that $\lim_{t\to\infty} P(T > t) = 0$. Let $\underline{T} = (T_1, \ldots, T_r)$ be the vector of component life lengths in an r-component system and let T denote the life length of the system. Let $\mathcal{I}$ denote the collection of nonempty subsets of $\{1, \ldots, r\}$. For each $I \in \mathcal{I}$, we say that <u>failure pattern I occurs</u> if the simultaneous failure of the components exclusively in subset I coincides with the failure of the system. Define

$$\xi(\underline{T}) = \begin{cases} I, & \text{if failure pattern I occurs} \\ \phi, & \text{otherwise.} \end{cases}$$

Let $\underline{S}$ and $\underline{T}$ represent the vectors of component life lengths of two systems whose system life lengths are S and T, respectively. We say that the two systems are <u>equivalent in life length and failure patterns</u> ($\underline{S} \overset{LP}{=} \underline{T}$, in symbols) if

$$P(S > t, \xi(\underline{S}) = I) = P(T > t, \xi(\underline{T}) = I), \; t \geq 0,$$

for every $I \in \mathcal{I}$.

The problem can now be stated as follows. Given the vector $\underline{T} = (T_1, \ldots, T_r)$ of (possibly dependent) life lengths, determine a random vector $\underline{S} = (S_1, \ldots, S_r)$

such that $\underline{S} \overset{LP}{=} \underline{T}$, where $S_1, \ldots, S_r$ are expressible in terms of independent random variables. The solution is found by letting each S_i be the life length of a component in a theoretical system, where the components are exposed to independent sources of shock as follows. Each component fails if it receives a shock. There exists one source of shock (sometimes referred to as a "hammerman") for each $I \in \mathcal{I}$. A shock from source I simultaneously kills the components exclusively in subset I. Let H_I denote the time (measured from the origin) until a shock from source I occurs. Then $S_i = \min(H_I, i \in I)$, $1 \leq i \leq r$, and $S = H$, where $S = \min(S_i, 1 \leq i \leq r)$ and $H = \min(H_I, I \in \mathcal{I})$. If the random vector $\underline{T}$ (and thus, $\underline{S}$) of life lengths has dimension r, then the vector $\underline{H}$ of independent times until shock has dimension $2^r - 1$. The model thus allows for simultaneous failures among the components in the original system. Define

$$\xi^*(\underline{H}) = \begin{cases} I, & \text{if } H_I < H_J \text{ for each } J \neq I \\ \phi, & \text{otherwise.} \end{cases}$$

It follows that $\underline{S} \overset{LP}{=} \underline{T}$ if and only if

$$P(H > t, \xi^*(\underline{H}) = I) = P(T > t, \xi(\underline{T}) = I) \qquad (2.1)$$

for each $t \geq 0$ and each $I \in \mathcal{I}$. If (2.1) holds for every subset I of $\{1, \ldots, r\}$, we write $\underline{H} \overset{LP}{=} \underline{T}$. We shall use the following notation throughout. For every life length T with distribution function F, let $\overline{F}(t) = P(T > t)$ denote the corresponding survival probability and let $\alpha(F) = \sup\{t: \overline{F}(t) > 0\}$. LPQ (1978) prove the following:

<u>Theorem 2.1</u>. Let $T = \min(T_i, 1 \le i \le r)$ denote the life length of an r-component series system, where T_i represents the life length of component i, i = 1, ..., r. Define $\overline{F}(t, I) = P(T > t, \xi(\underline{T}) = I)$ and $F(t, I) = P(T \le t, \xi(\underline{T}) = I)$, $I \in \mathcal{I}$.

Then the following statements hold:

(i) A necessary and sufficient condition for the existence of a set of independent random variables $\{H_I, I \in \mathcal{I}\}$ which satisfy $\underline{H} \overset{LP}{=} \underline{T}$, where $H = \min(H_I, I \in \mathcal{I})$, is that the functions $F(\cdot, I)$, $I \in \mathcal{I}$, have no common discontinuities in the interval $[0, \alpha(F))$.

(ii) The random variables $\{H_I, I \in \mathcal{I}\}$ in (i) have corresponding survival probabilities $\{\overline{G}_I(\cdot), I \in \mathcal{I}\}$ which are uniquely determined on the interval $[0, \alpha(F)]$ as follows:

$$\overline{G}_I(t) = \prod_{a \le t} [\overline{F}(a)/\overline{F}(a^-)] \exp[-\int_0^t (dF^C(\cdot, I)/\overline{F}), \quad 0 \le t \le \alpha(F), \tag{2.2}$$

where $F^C(\cdot, I)$ is the continuous part of $F(\cdot, I)$, the product is over the set of discontinuities $\{a\}$ of $F(\cdot, I)$, $I \in \mathcal{I}$, and the product over an empty set is defined as unity.

<u>Remark 2.2</u>. Formula (2.2) is defined in LPQ (1978) for t in the half-open interval $[0, \alpha(F))$. The present formulation is, however, equivalent.

<u>Remark 2.3</u>. In Theorem 2.1, suppose that failure pattern I is non-occurring for some $I \in \mathcal{I}$, i.e., suppose $P(\xi(\underline{T}) = I) = 0$. Then $F(t, I) = P(T \le t, \xi(\underline{T}) = I) = 0$ for each $t \ge 0$. By (2.2), $\overline{G}_I \equiv 1$ on the interval $[0, \alpha(F)]$. The corresponding random variable H_I has a distribution which places no mass on the interval $[0, \alpha(F)]$. We can view H_I as corresponding to a hammerman who does not strike in the interval $[0, \alpha(F)]$. If $\alpha(F) = \infty$, then H_I is degenerate

at infinity (and so is not a life length). In the same way, every non-occurring pattern can be associated with a hammerman who does not strike. It follows from Theorem 2.1 that if the original system has exactly m occurring failure patterns, $1 \le m \le 2^r - 1$, then the original system is equivalent in life length and failure patterns to a system involving the same number m of independent random variables $\{H_i, 1 \le i \le m\}$. In particular, suppose we add to the hypothesis of Theorem 2.1 the assumption that death (failure) does not result from simultaneous causes (so that $P(T_i = T_j) = 0$ for $i \ne j$). Then the original system has at most r occurring failure patterns and so can be replaced by a system involving at most r independent random variables H_i.

Remark 2.4. Suppose that the original vector $\underline{T}$ in Theorem 2.1 is itself a vector of independent random variables and the functions $F(\cdot, I)$, $I \in \mathcal{I}$, have no common discontinuities. Then it is not difficult to show that a solution $\underline{S}$ to the above problem is such that $\underline{S}$ and $\underline{T}$ have the same distribution.

3. The independent competing risks model: the Kaplan-Meier estimator.

Let there be a finite number of causes of death labelled 1, ..., r. We associate with each cause j a nonnegative random variable T_j, j = 1, ..., r. The random variable T_j represents the age at death if cause j were the only cause present in the environment. [In a reliability setting T_j denotes the life length of component j in a series system of r components. In an incomplete or censored data problem, one of the random variables T_j represents the time at which an individual becomes "unobservable" for a reason other than death, while the remaining variables typically represent various causes of death. The complete collection of random variables $T_1, \ldots, T_r$ is not observed. Instead, only two quantities are observed: the age at death given by $\tau = \min(T_1, \ldots, T_r)$ and the cause of death, labelled ξ, given by the subset I of $\{1, \ldots, r\}$ such that $\tau = T_i$ for each $i \in I$ and $\tau \neq T_i$ for each $i \notin I$. When death results from exactly one of the r possible causes, as is usually assumed, then ξ is the index i for which $\tau = T_i$. The biomedical researcher is interested in making inferences about unobservable quantities (viz., the random variables $T_1, \ldots, T_r$) by using data from observable quantities--in this case, the lifetime τ and the cause of death ξ. In particular, he seeks to estimate the marginal survival probability corresponding to a given cause (or combination of causes) operating along without competition from the other causes. That is, he wishes to estimate the $2^r - 1$ survival probabilities

$$\overline{M}_J(t) = P[\min(T_j,\ j \in J) > t],\ J \subset \{1, \ldots, r\}.$$

In analyzing competing risk data, various authors typically assume one or more of the following:

(A1) The risks (i.e., the random variables $T_1, \ldots, T_r$) are mutually independent.

(A2) Death does not result from simultaneous causes. [Consequently, $P(T_i = T_j) = 0$ for $i \neq j$.]

(A3) The distributions of $T_1, \ldots, T_r$ have no common discontinuities.

(A4) The random variables $T_1, \ldots, T_r$ have a joint distribution which is absolutely continuous.

For many years there have been several approaches to problems of estimation in competing risk theory which employ, in varying degrees, the above assumptions. For example, the assumption of independence (A1) was until recently almost universally made even though it is obviously inappropriate in many problems. Moreover, assumptions (A2) through (A4) need not hold in certain situations of interest. For example, (A2) and (A4) do not hold in engineering systems where system failure can occur as a result of the simultaneous failures of two (or more) components. Assuming (A1), (A2) and (A3), Peterson (1975, 1977) shows how the Kaplan-Meier estimator may be expressed as a function of the empirical counterparts of the functions $F(\cdot, I)$, $I \in \mathcal{I}$, in Theorem 2.1. He thus indicates a way to obtain strong consistency of the estimator when $\underline{T}$ is a vector of discrete random variables. In this paper we show how Theorem 2.1 can be used to approach the problem of estimating the marginal distributions of interest in a unified way without making <u>any</u> of the assumptions (A2), (A3), or (A4). We assume only a weaker version of (A3). Moreover, we are able to drop the assumption of independence (A1) and find necessary and sufficient conditions for the existence of consistent estimators for the marginal distributions of interest. Peterson (1975) also considers the case of dependent risks. In Section 4 we show how our methods extend and simplify those of Peterson.

Let $\underline{T}_i = (T_{1i}, \ldots, T_{ri})$, $i = 1, \ldots, n$, represent a random sample from the joint distribution of the nonnegative random variables $T_1, \ldots, T_r$. Denote the marginal distributions (survival probabilities) of $T_1, \ldots, T_r$ by $M_i(\overline{M}_i)$, $i = 1, \ldots, r$. For each $I \in \mathcal{I}$, let $M_I(t) = P(T_I \leq t)$, where $T_I = \min(T_i, i \in I)$. Assume (A1), (A2), and (A3). Then the cause of death $\xi(\underline{T}) = i$ if and only if $T_i < T_j$ for each $j \neq i$, $1 \leq i, j \leq r$. For each $i = 1, \ldots, n$, only τ_i and ξ_i are observed, where $\tau_i = \min(T_{1i}, \ldots, T_{ri})$ and $\xi_i = j$ whenever $\tau_i = T_{ji}$. Consider the case $r = 2$ and suppose we seek to estimate the marginal survival probability $M_1(t) = P(T_1 > t)$. Let $0 \equiv \tau_{(0)} \leq \tau_{(1)} \leq \ldots \leq \tau_{(n)}$ denote the ordered values of the observations $\tau_1, \ldots, \tau_n$. Then the Kaplan-Meier (K-M) estimator of $\overline{M}_1(t)$ is

$$\hat{\overline{M}}_1(t) = \prod_i [(n - i)/(n - i + 1)], \tag{3.1}$$

where the product is over the ranks i of those ordered observations $\tau_{(i)}$, $1 \leq i \leq n$, such that $\tau_{(i)} \leq t < \tau_{(n)}$ and $\tau_{(i)}$ corresponds to a death from cause 1 [$\tau_{(i)} = T_{1j}$ for some j]. If $\tau_{(n)}$ corresponds to a death from cause 1, then $\hat{\overline{M}}_1(t)$ is defined to be zero for $t > \tau_{(n)}$. Otherwise, $\hat{\overline{M}}_1(t)$ is undefined for $t > \tau_{(n)}$. [In the original formulation by Kaplan and Meier (1958), T_1 corresponded to the time until death, while T_2 represented the time at which a loss occurred.]

The K-M estimator (3.1) is a step function with jumps at those observations τ_i which correspond to a death from cause 1. If no death from cause 2 is observed [no loss occurs], then each observation corresponds to a death from cause 1. In this case (3.1) reduces to a step function with jumps of height $1/n$ at each τ_i, thus yielding the usual empirical estimate of $\overline{M}_1(t)$.

Peterson (1975) considers the following straightforward extension of (3.1) for the survival probabilities $\overline{M}_J(t) = P(T_J > t)$, $J \subset \{1, \ldots, r\}$, where $T_J = \min(T_j, j \in J)$ and $r \geq 2$. The (extended) K-M estimator $\hat{\overline{M}}_J$ is given by

$$\hat{\overline{M}}_J(t) = \prod_i [(n - i)/(n - i + 1)], \tag{3.2}$$

where now the product is over the ranks i of those ordered observations $\tau_{(i)}$ such that $\tau_{(i)} \leq t < \tau_{(n)}$ and $\tau_{(i)}$ corresponds to a death from <u>at least one</u> cause $j \in J$. Conventions analogous to those used in defining (3.1) also hold in (3.2) when $t > \tau_{(n)}$.

Assuming independence (A1), no simultaneous causes of death (A2), and disjoint sets of discontinuities for the marginal distributions (A3), Peterson (1977) indicates a way to obtain strong consistency of the K-M estimator (3.1) in the special case when the random variables $T_1, \ldots, T_r$ are discrete. Breslow and Crowley (1974) etc. Breslow and Crowley (1974) and Meier (1975) show that the estimator is asymptotically normal and, as a process in t, converges to a normal process. More recently, Aalen (1976) shows that the bivariate vector of K-M estimators $(1 - \hat{\overline{M}}_1(t_1), 1 - \hat{\overline{M}}_2(t_2))$ is asymptotically bivariate normal, and that, regarded as a bivariate process in t_1 and t_2, converges to a normal process. Estimators analogous to (3.1) and (3.2) are proposed in the next section by using formula (2.2) of Theorem 2.1. Such extensions will apply in situations <u>where the assumptions of independence (A1) and no simultaneous causes of death (A2) fail to hold.</u>

4. The dependent case.

In this section we show how formula (2.2) can be used to unify and simplify the nonparametric approach toward estimation in the competing risks model when independence of risks is not assumed. First, we show how (2.2) easily establishes a functional relationship between distributions of "theoretical" random variables [the $\{H_I, I \in \mathcal{I}\}$ of Theorem 2.1] which are unobservable (but estimable) and the marginal distributions M_i, $1 \le i \le r$, (of unobservable variables) for which we seek estimates. Under a weaker version of assumption (A3), consistent estimators analogous to (3.2) are given.

In the probabilistic problem discussed in Section 2, our solution (Theorem 2.1) can be described as follows. Replace the given series system of dependent components by a "theoretical" system whose component life lengths are expressible in terms of independent random variables so as to preserve the joint distribution of system life length and failure patterns. The components in the so-called theoretical system are exposed to shocks from "hammermen" whose striking times H_I, $I \in \mathcal{I}$, have distributions defined by (2.2). A series system of r components with dependent life lengths $T_1, \ldots, T_r$, and system life length $T = \min(T_i, 1 \le i \le r)$ becomes, in the terminology of competing risks, an individual exposed to r dependent risks of death, where T_i represents the age at death if risk i were the only risk present in the environment, $i = 1, \ldots, r$. In this section, unless otherwise indicated, we drop the assumption (A1) of independent risks. With respect to the estimation problem posed in Section 3, formula (2.2) exhibits a relationship between distributions of observable quantities and distributions associated with theoretical variables H_I, $I \in \mathcal{I}$, which are unobservable. The observable quantities in the competing risks model are the life length $\tau = \min(T_i, 1 \leq i \leq r)$ and the cause of death ξ. The functions associated with these observable quantities, viz., the survival probability $\overline{F}(t) = P(T > t)$ and the monotonic functions $F(t, i) = P(T \leq t, \xi(T) = i)$, $1 \leq i \leq r$, appear on the right hand side of (2.2). Thus replacing $\overline{F}(t)$ and $\overline{F}(t, i)$ in (2.2) with their empirical counterparts allows us to <u>estimate</u> the distributions G_i, $1 \leq i \leq r$, associated with the unobservable variables H_i, $1 \leq i \leq r$. However, the distributions G_i, $1 \leq i \leq r$, are, in general, <u>different</u> from the marginal distributions M_i, $1 \leq i \leq r$, which we seek to estimate. The natural question then is how to relate the unobservable (but estimable) functions G_i, $1 \leq i \leq r$, to the marginal distributions M_i, $1 \leq i \leq r$. More generally, how can we relate the functions M_I, $I \in \mathcal{I}$, to the survival probabilities $\overline{G}_I$, $I \in \mathcal{I}$, given by (2.2)? One answer can be given as follows. That is, suppose for a moment that $T_1, \ldots, T_r$ are, in fact, <u>independent</u> and that the following holds:

(A3)' The functions $F(\cdot, I)$, $I \in \mathcal{I}$, in Theorem 2.1 have no common discontinuities.

Then by Remark 2.4, $G_i = M_i$, $i = 1, \ldots, r$. Consequently,

$$\overline{M}_I(t) = \prod_{i \in I} \overline{M}_i(t) = \prod_{i \in I} \overline{G}_i(t) \tag{4.1}$$

for every $t \in [0, \alpha(F)]$. What are the corresponding estimators? It is a simple exercise to show that in the case $r = 2$, if we replace the functions $\overline{F}(t)$ and $\overline{F}(t, 1)$ on the right in (2.2) by their empirical counterparts, then the resulting statistic is <u>the K-M estimator (3.1)</u> of $P(T_1 > t)$. In view of (4.1), if r is an arbitrary integer greater than 2, a reasonable estimator for $\overline{M}_I(t)$ ought to be $\Pi\, \hat{\overline{G}}_i(t)$, where the product is over $i \in I$ and $\hat{\overline{G}}_i$ is the function resulting from (2.2) by replacing the functions $\overline{F}(\cdot)$ and $\overline{F}(\cdot, i)$ with their empirical counterparts, $i = 1, \ldots, r$. Again, it is easy to show that the resulting statistic is, in fact, the (generalized) K-M estimator (3.2) for $\overline{M}_I(t)$. Thus, in the case of independent risks, (2.2) leads directly to well-known estimators possessing several optimal properties. It is therefore reasonable to expect that (2.2) also plays a role in the estimation problem when the risks are mutually dependent. This is, indeed, the case.

Theorem 4.1. [Peterson (1975)]. Let $T_1, \ldots, T_r$ be nonnegative random variables satisfying (A2) and (A3) [but not necessarily (A1)]. Let $\mathcal{P}$ be a partition of $\{1, \ldots, r\}$ and define $\overline{G}_i$, $1 \le i \le r$, as in (2.2). Then for each $t \in [0, \alpha(F)]$,

$$\overline{M}_I(t) = \Pi_{i \in I}\, \overline{G}_i(t) \ \underline{\text{for each } I \in \mathcal{P}} \tag{4.2a}$$

if and only if

$$P(T > t, \xi(\underline{T}) \in I) = \int_t^\infty \overline{M}_{I'}(x)\, dM_I(x) \ \underline{\text{for each } I \in \mathcal{P}}, \tag{4.2b}$$

where I' denotes the complement of I in $\{1, \ldots, r\}$.

Peterson (1975) uses an operator defined on a space of distribution functions to prove an equivalent version of Theorem 4.1. Employing Theorem 2.1, we give a proof which is considerably simpler. The following notation and lemma is useful in interpreting (4.2a, b).

Let $\{T_i, 1 \le i \le r\}$ and $\{T_i^*, 1 \le i \le r\}$ be two collections of random variables. For each function f of the random variables $T_1, \ldots, T_r$, let f^* denote the value of the same function of $T_1^*, \ldots, T_r^*$. For each set I belonging to a partition $\mathcal{P}$ of $\{1, \ldots, r\}$, define $\overline{F}_{\mathcal{P}}(t, I) = P(T > t, \xi(\underline{T}) \in I)$. On the right in (2.2), replace $F(t, I)$ by $F_{\mathcal{P}}(t, I)$ and call the resulting expression $\overline{G}_{I,\mathcal{P}}(t)$.

The following simple lemma is useful in proving Theorem 4.1:

Lemma 4.2. Assume the hypothesis of Theorem 4.1. Let $T_1^*, \ldots, T_r^*$ be independent random variables such that T_I^* and T_I have the same distribution (i.e., $M_I^* = M_I$) for each $I \in P$. Then (4.2a) is equivalent to

$$\overline{G}^*_{I,P} = \overline{G}_{I,P} \text{ for each } I \in P, \tag{4.3a}$$

and (4.2b) is equivalent to

$$\overline{F}_P(t, I) = \overline{F}_P^*(t, I) \text{ for each } I \in P. \tag{4.3b}$$

Proof. By (A2) and (A3), we have that M_I and M_J have no common discontinuities and $P(T_I = T_J) = 0$ for each $I, J \in P$, $I \neq J$. Since $M_I = M_I^*$, we deduce that $P(T_I^* = T_J^*) = 0$ for each $I, J \in P$, $I \neq J$. It follows from Remark 2.3 that if P has exactly k members, $2 \leq k \leq r$, then each of the collections $\{T_I, I \in P\}$ and $\{T_I^*, I \in P\}$ has at most k occurring failure patterns. By Remark 2.4,

$$\overline{M}_I^* = \overline{G}_{I,P} \; (= \overline{M}_I). \tag{4.4}$$

By (2.2) and the fact that $\overline{F}_P(t, I) = \sum_{i \in I} \overline{F}(t, i)$,

$$\overline{G}_{I,P} = \Pi_{i \in I} \overline{G}_i. \tag{4.5}$$

A simple calculation shows that

$$\overline{F}_P^*(t, I) = \int_t^\infty \overline{M}_{I'}(x)\, dM_I(x). \tag{4.6}$$

The conclusion follows from (4.4), (4.5), and (4.6). ||

We now give the following elementary proof of Theorem 4.1.

<u>Proof of Theorem 4.1</u>. By Lemma 4.2, it is enough to show that (4.3a) holds if and only if (4.3b) holds.

Suppose first that (4.3b) holds. Since

$$\overline{F}(t) = \sum_{I\varepsilon P} \overline{F}_P(t, I) \text{ and } \overline{F}^*(t) = \sum_{I\varepsilon P} \overline{F}^*(t, I),$$

it follows from (2.2) that $\overline{G}_{I,P} = \overline{G}^*_{I,P}$.

Conversely, suppose that (4.3a) holds. It follows from (2.2) that $\overline{F} = \Pi\, \overline{G}_{I,P}$ and $\overline{F}^* = \Pi\, \overline{G}^*_{I,P}$, where each product is over $I \varepsilon P$. Since

$$\overline{F}_P(t, I) = \int_t^\infty (\overline{F}/\overline{G}_{I,P})dG_{I,P}$$

and

$$\overline{F}_P^*(t, I) = \int_t^\infty (\overline{F}^*/\overline{G}^*_{I,P})\, dG^*_{I,P}$$

for each $I \varepsilon P$, relation (4.3b) holds. ||

One drawback of Peterson's formulation is that the assumption of no simultaneous causes of death (A2) does not hold, e.g., in an engineering system where system failure can occur as a result of the simultaneous failure of two or more components. Moreover, Peterson (1975, 1977) assumes that the <u>marginal</u> distributions of $T_1, \ldots, T_r$ have no common discontinuities (A3). In Theorem 4.4 below we give necessary and sufficient conditions for a functional relation to exist between the functions $\overline{G}_i$ in (2.2) and the functions M_I which we seek to estimate <u>without assuming (A2)</u>. Our only assumption is that the functions $F(\cdot, I)$, $I \varepsilon \mathcal{I}$, have no common discontinuities in $[0, \alpha(F)]$ [assumption (A3)']. Assumption (A3) implies (A3)', but the converse does not hold.

Another disadvantage in Peterson's (1975) approach can be described as follows. In order to establish a relation between the function M_J for an individual subset J of {1, ..., r} and the functions $\overline{G}_I$, $I \in \mathcal{I}$, in (2.2), Peterson (1975) requires that (4.2b) hold simultaneously for each set I in a partition $\mathcal{P}$ of {1, ..., r} which contains J. In general, it is not difficult to construct joint distributions which satisfy none of the conditions in (4.2b) yet for which at least one relationship exists between the function M_J and the survival probabilities $\overline{G}_I$, $I \in \mathcal{I}$.

Example 4.3. Let the discrete random vector (T_1, T_2) have a joint probability distribution as given in Table 4.1 below:

Table 4.1. Distribution of (T_1, T_2)

$T_1 \backslash T_2$	2	4	6
1	1/24	1/8	1/12
3	1/12	1/16	0
5	1/6	1/6	1/6

A simple calculation verifies that $\overline{G}_1$ given by (2.2) equals $\overline{M}_1$, but $\overline{G}_2 \neq \overline{M}_2$. Furthermore, neither of the conditions (4.2b) is satisfied. Thus, Theorem 4.1 above is not applicable in this simple example. We will show, however, that our generalization of Theorem 4.1 (Theorem 4.4 below) does apply here. Before we state Theorem 4.4, it will be conveneint to introduce the following notation. Let $\mathcal{I}_I = \{J \in \mathcal{I}:\ J \cap I \neq \phi\}$. Let $F(t, \mathcal{I}_I) = P(T > t, \xi(\underline{T}) \in \mathcal{I}_I)$. Recall that the cause of death is subset I [i.e., $\xi(\underline{T}) = I$] if and only if $T = T_i$ for each $i \in I$ and $T \neq T_i$ for each $i \notin I$. Thus, $F(t, \mathcal{I}_I) = \sum P(T > t, \xi(\underline{T}) = J) = \sum F(t, J)$, where each sum is over $J \in \mathcal{I}_I$. For each function G, let D(G) [C(G)] denote the set of discontinuities (continuities) of G. Note that if the functions $F(\cdot, I)$, $I \in \mathcal{I}$, have no common discontinuities, then $D[F(\cdot, \mathcal{I}_I)] = \cup D[F(\cdot, J)]$, where the union is disjoint over sets $J \in \mathcal{I}_I$.

Theorem 4.4 below resembles Theorem 4.1 in that we find necessary and sufficient conditions for a relationship to exist between the functions M_I, $I \in \mathcal{I}$, and the survival probabilities $\overline{G}_I$, $I \in \mathcal{I}$, given by (2.2). It generalizes Theorem 4.1 in the following ways. First, the assumption (A2) of no simultaneous causes of death is dropped. Secondly, we assume a weaker version of assumption (A3), namely the assumption (A3)' that the functions $F(\cdot, I)$, $I \in \mathcal{I}$, have no common discontinuities.

Theorem 4.4. Let $T_1, \ldots, T_r$ be nonnegative random variables satisfying (A3)'. Let $I \in \mathcal{I}$. Then for each $t \in [0, \alpha(F)]$,

$$\overline{M}_I(t) = \prod_{J \in \mathcal{I}_I} \overline{G}_J(t) \tag{4.7}$$

if and only if the following two conditions hold:

$$M_I(a)/M_I(a^-) = \begin{cases} \overline{F}(a)/\overline{F}(a^-), & a \in D(F(\cdot, \mathcal{I}_I)) \\ 1, & \text{otherwise.} \end{cases} \tag{4.8a}$$

and

$$P(T_{I'} \geq t \mid T_I = t) = P(T_{I'} > t \mid T_I > t), \tag{4.8b}$$

where $\overline{G}_J$ is given by (2.2).

The proof of Theorem 4.4 is given in Section 5.

Remark 4.5. Suppose that the random variables $T_I = \min(T_i, i \in I)$, $I \in \mathcal{I}$, have absolutely continuous distributions. Let $m_I(t)$ [$M_I(t)$] and $m_{I|I'}(t)$ [$M_{I|I'}(t)$] denote respectively the density (distribution) function and conditional density (distribution) function of T_I and T_I given $T_{I'} > t$. Then condition (4.8b) is equivalent to

$$m_{I|I'}(t)/M_{I|I'}(t) = m_I(t)/M_I(t).$$

In other words, the conditional failure rate function of T_I given $T_{I'} > t$ is equal to the (unconditional) failure rate function of T_I. Stated differently, the random variables T_I and $T_{I'}$ are independent "along the diagonal $T_I = T_{I'}$". Desu and Narula (1977) arrive at a condition similar to (4.8b) in the special case when the assumption of absolute continuity (A4) [and hence also (A2)] holds.

Note that conditions (4.8a, b), in contrast to (4.2b), apply to only one subset $I \in \mathcal{I}$ at a time. Consequently, we can proceed in Example 4.3 as follows. It is easy to verify that conditions (4.8a, b) hold. Since $\overline{G}_{\{1,2\}} \equiv 1$, it follows from Theorem 4.4 that $\overline{M}_1 = \overline{G}_1$.

We have previously remarked that the assumption (A2) of no simultaneous causes of death is unrealistic in certain models of interest. An important family of multivariate distributions for which assumption (A2) fails is the family of multivariate exponential (MVE) distributions of Marshall and Olkin (1967). We illustrate with an example.

Example 4.6. For simplicity, suppose that the random vector (T_1, T_2) has the Marshall-Olkin bivariate exponential distribution with survival probability:

$$P(T_1 > t_1, T_2 > t_2) = \exp[-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)],$$

for t_1, $t_2 \geq 0$ and $\lambda_1, \lambda_2, \lambda_{12} > 0$. Since the marginal distributions M_1 and M_2 are continuous, condition (4.8a) trivially holds. Condition (4.8b) with $I = \{1\}$ states that

$$P(T_2 \geq t | T_1 = t) = P(T_2 > t | T_1 > t).$$

An easy computation shows that these conditional probabilities are each equal to $\exp(-\lambda_2 t)$. Thus, Theorem 4.4 may be applied when the joint distribution belongs to the family of Marshall-Olkin MVE distributions, whereas Theorem 4.1 cannot be applied here since (A2) fails.

In Section 3 we assumed that the risks were mutually independent (A1)' and that the functions $F(\cdot, I)$, $I \in \mathcal{I}$, had no common discontinuities (A3)'. Under these assumptions the basic formula (2.2) yielded the K-M estimators for the marginal distribution M_1 $(r = 2)$ and the functions M_I, $I \in \mathcal{I}$, $(r \geq 2)$. In a similar fashion formula (2.2) (via Theorem 4.4) can be used to determine strongly consistent estimators for the functions M_I, $I \in \mathcal{I}$, in the important practical cases when independence fails to hold and simultaneous causes of death are allowed.

The key tool we shall use in establishing consistent estimators for the marginal distributions of interest is given in Theorem 4.7 below. First we introduce some notation. As above, $\underset{\sim}{T} = (T_1, \ldots, T_r)$ is a vector of nonnegative random variables, $\tau = \min(T_j,\ 1 \le j \le r)$ represents the life length of an individual exposed to r risks of death and ξ represents the cause of death, where $\xi = J$ if and only if $\tau = T_j$ for each $j \in J$ and $\tau \ne T_j$ for each $j \notin J$, $J \in \mathcal{I}$. For each $I \in \mathcal{I}$ and Borel set A, let $F(A, I) = P(\tau \in A, \xi = I)$. Let τ have distribution function F and let D(I) [C(I)] be the set of discontinuities [continuities] of the function $F(\cdot, I)$. There exists, by Theorem 2.1, a collection $\{H_I, I \in \mathcal{I}\}$ of independent random variables such that $\underline{H} \overset{LP}{=} \underline{T}$, where $H = \min(H_I, I \in \mathcal{I})$. Moreover, the probability $\overline{G}_I(t^+) = P(H_I \ge t)$ may be obtained from (2.2). Now let $\underline{T}_i = (T_{1i}, \ldots, T_{ri})$, i = 1, 2, ..., be a sequence of nonnegative random vectors (representing a sequence of individuals) and let τ_i, ξ_i, $F_i(\cdot, I)$, F_i, D(I, i), C(I, i), $H_{I,i}$, H_i, and $\overline{G}_{I,i}$ be the analogues of τ, ξ, $F(\cdot, I)$, F, D(I), C(I), H_I, H, and $\overline{G}_I$ above. Let $F^C(F^D)$ denote the continuous (discontinuous) part of F.

Theorem 4.7. Suppose the following conditions hold:

4(i) For $I \ne J$, the pair $\{F(\cdot, I), F(\cdot, J)\}$ as well as each pair $\{F_i(\cdot, I), F_i(\cdot, J)\}$, i = 1, 2, ..., have no common discontinuities.

4(ii) For $I \in \mathcal{I}$ and $0 \le t \le \alpha(F)$,

$$\lim_{n\to\infty} F_n(\{[0, t] \cap C(I)\}, I) = F^C([0, t], I).$$

4(iii) For $I \in \mathcal{I}$,

$$\lim_{n\to\infty} \sup_{0\le t\le\alpha(F)} |F_n(([0, u] \cap D(I)), I) - F^D([0, u], I)| = 0.$$

4(iv) $\lim_{n\to\infty} \sup_{0\le t\le\alpha(F)} |\overline{F}_n(t) - \overline{F}(t)| = 0.$

Then for $I \in \mathcal{I}$ and $0 \le t \le \alpha(F)$,

$$\lim_{n\to\infty} G_{I,n}(t^+) = G_I(t^+). \tag{4.9}$$

The proof of Theorem 4.7 is given in Section 5.

Consider now the estimation problem posed above. Let $\underline{T}_i = (T_{1i}, \ldots, T_{ri})$, $i = 1, 2, \ldots$, be independent and identically distributed as $\underline{T} = (T_1, \ldots, T_r)$. Replace $F(\cdot, I)$ and F by their empirical counterparts, $\hat{F}_n(\cdot, I)$ and $\hat{F}_n$, respectively, on the right in (2.2). The resulting expression, call it $\hat{\overline{G}}_{I,n}$, is an estimator for $\overline{G}_I$. Assume that $\underline{T}$ satisfies (4.8a, b). Then (4.7) holds. A natural estimator for $M_I(t)$ in this case, then, is

$$\hat{\overline{M}}_{I,n}(t) = \prod_{J\in\mathcal{I}_I} \hat{\overline{G}}_{J,n}(t). \tag{4.10}$$

Such an estimator will be strongly consistent if for every $J \in \mathcal{I}_I$,

$$\hat{\overline{G}}_{J,n}(t^+) \to \overline{G}_J(t^+) \text{ a.s.} \tag{4.11}$$

We now show that this is a simple consequence of Theorem 4.7.

In Theorem 4.7 identify $F_n(\cdot, I)$ with $\hat{F}_n(\cdot, I)$ and F_n with $\hat{F}_n$, where

$$\hat{F}_n(t, I) = n^{-1} \sum_{i=1}^{n} \chi_{\{\tau_i \le t,\ \xi_i = I\}}$$

$$\hat{F}_n(t) = n^{-1} \sum_{i=1}^{n} \chi_{\{\tau_i \le t\}},$$

and χ_A is the indicator function of the set A. Define

$$\hat{F}_{C(I),n}(t) = n^{-1} \sum_{i=1}^{n} \chi\{\tau_i \in [0, t] \cap C(I), \xi_i = I\}$$

$$\hat{F}_{D(I),n}(t) \quad n^{-1} \sum_{i=1}^{n} \chi\{\tau_i \in [0, t] \cap D(I), \xi_i = I\}.$$

It follows that

$$\hat{F}_n(\cdot, I) = \hat{F}_{C(I),n} + \hat{F}_{D(I),n} \tag{4.12}$$

and

$$\hat{F}_n = \sum_{I \in \mathcal{I}} \hat{F}_{I,n}. \tag{4.13}$$

We denote the common domain of all random variables by Ω. To prove (4.11), it suffices to verify conditions 4(i) - 4(iv) of Theorem 4.7 for almost every $\omega \in \Omega$.

Condition 4(i) holds trivially. By the strong law of large numbers, we have that as $n \to \infty$,

$$\hat{F}_{C(I),n}(t) \to F^C([0, t], I) \text{ a.s., } t \geq 0 \tag{4.14}$$

and

$$\hat{F}_{D(I),n}(t) \to F^D([0, t], I) \text{ a.s., } t \geq 0. \tag{4.15}$$

Condition 4(ii) follows from (4.14). By the Glivenko-Cantelli Theorem,

$$\lim_{n\to\infty} \sup_{t \geq 0} |\hat{F}_n(t) - F(t)| \to 0 \text{ a.s.} \tag{4.16}$$

Thus, condition 4(iv) holds. Since the function $F^C([0, t], I)$ is a continuous function of t, it follows that the convergence in (4.14) is uniform in t. Thus, we may conclude from (4.12) and (4.16) that the convergence in (4.15) is uniform in $t \geq 0$ and so condition 4(iii) holds.

Remark 4.8. Let $0 \equiv \tau_{(0)} \leq \tau_{(1)} \leq \ldots \leq \tau_{(n)}$ denote the ordered values of $\tau_1, \ldots, \tau_n$. In analogy with the Kaplan-Meier estimator (3.2), the estimator (4.10) may be expressed as follows:

$$\hat{\bar{M}}_I(t) = \prod_i [(n - i)/(n - i + 1)], \tag{4.17}$$

where the product is over the ranks i of those ordered observations $\tau_{(i)}$, $1 \leq i \leq n$, such that $\tau_{(i)} \leq t < \tau_{(n)}$ and $\tau_{(i)}$ corresponds to a death from the simultaneous causes $j \in J$, $J \in \mathcal{I}_I$. If for some i, $\tau_{(n)} = T_{ji}$ for each $j \in J$, $J \in \mathcal{I}_I$, then (4.17) is defined to be zero for $t > \tau_{(n)}$. Otherwise, (4.17) is undefined for $t > \tau_{(n)}$.

In view of Remark 4.8 and the preceding argument, we have proven:

Theorem 4.9. In the competing risks model of Section 3, assume only that the functions $F(\cdot, I)$, $I \in \mathcal{I}$, have no common discontinuities, and the joint distribution of $(T_1, \ldots, T_r)$ satisfies (4.8a, b). Then the estimator (4.17) is strongly consistent for $\bar{M}_I$.

Remark 4.10. To show consistency of the K-M estimator in the independent case, Peterson (1977) must rely on a property of an operator defined on a space of discrete distribution functions (which he states without proof). If we assume that the risks are independent and that assumption (A3)' holds, then by applying Theorem 4.7 above, as we did in the dependent case, we not only have a stronger proof of the consistency of the K-M estimator [since (A3) implies (A3)'], but also one which is considerably more elementary.

5. Proofs.

Before we give a proof of Theorem 4.4, we state two lemmas which are proven in LPQ (1978).

Lemma 5.1. For every probability measure Q (with a possible atom at ∞) such that $\overline{Q}(0^-) = 1$, and every $t \geq 0$, the following holds:

$$\overline{Q}(t) = \exp[-\int_0^t (dQ^C/\overline{Q}) \cdot \prod_{a \leq t} [\overline{Q}(a)/\overline{Q}(a^-)], \tag{5.1}$$

where the product is over the set $\{a\}$ of discontinuities of Q, and the product over an empty set is defined to be 1.

Lemma 5.2. Let $\{\overline{G}_I, I \in \mathcal{I}\}$ be a collection of survival probabilities satisfying (2.2). Then for each $I \in \mathcal{I}$ and $t \in [0, \alpha(F))$,

$$\overline{G}_I(t)/\overline{G}_I(t^-) = \begin{cases} \overline{F}(t)/\overline{F}(t^-), & t \in D(F(\cdot, I)) \\ 1 & \text{otherwise.} \end{cases}$$

Proof of Theorem 4.4. Suppose (4.8a, b) holds. By (4.8a),

$$D(F(\cdot, \mathcal{I}_I)) = D(M_I), \; I \in \mathcal{I}. \tag{5.2}$$

For every Borel set $B \subset [0, \alpha(F))$,

$$\begin{aligned} F(B, \mathcal{I}_I) &\equiv P(T \in B, \xi(\underline{T}) \in \mathcal{I}_I) \\ &= P(T_I \in B, T_I \leq T_{I'}) \\ &= \int_B P(T_{I'} \geq u | T_I = u) \, dM_I(u) \\ &= \int_B P(T_{I'} > u | T_I > u) \, dM_I(u) \; [\text{by } 4.8(b)] \\ &= \int_B [\overline{F}(u)/\overline{M}_I(u)] \, dM_I(u). \end{aligned}$$

Thus, for every Borel set $B \subset [0, \alpha(F))$,

$$F(B, I_I) = \int_B (\overline{F}/\overline{M}_I)\, dM_I. \tag{5.3}$$

Relations (5.2) and (5.3) together imply that

$$dF^C(\cdot, I_I)/dM_I^{\,C} = \overline{F}/M_I.$$

It follows that

$$\prod_{J \in I_I} \overline{G}_J(t) = \exp\{-\int_0^t [dF^C(\cdot, I_I)/\overline{F}]\} \prod_{a \in D(F(\cdot, I_I))} [\overline{F}(a)/\overline{F}(a^-)]$$

$$= \exp\{-\int_0^t [dM_I^{\,C}/\overline{M}_I]\} \prod_{a \in D(M_I)} [\overline{M}_I(a)/\overline{M}_I(a^-)]$$

[by 4.8a]

$$= \overline{M}_I(t) \text{ [by Lemma 5.1].}$$

Conversely, suppose (4.7) holds. By (2.2) and Lemma 5.1,

$$\exp\{-\int_0^t [dF^C(\cdot, I_I)/\overline{F}]\} \prod_{\substack{a \le t \\ a \in D(F(\cdot, I_I))}} [\overline{F}(a)/\overline{F}(a^-)] \tag{5.4}$$

$$= \exp[-\int_0^t (dM_I^{\,C}/\overline{M}_I)] \prod_{\substack{a \le t \\ a \in D(M_I)}} [\overline{M}_I(a)/\overline{M}_I(a^-)].$$

Letting Π denote the product over sets $J \in I_I$, we have

$$\overline{M}_I(a)(\overline{M}_I(a^-) = \Pi[\overline{G}_J(a)]/\Pi[\overline{G}_J(a^-)]$$

$$= \begin{cases} \Pi[\overline{F}(a)/\overline{F}(a^-)], & a \in D(F(\cdot, J)) \\ 1 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \overline{F}(a)/\overline{F}(a^-), & a \in D(F(\cdot, I_I)) \\ 1 & \text{otherwise.} \end{cases}$$

Thus, (4.8a) holds. Equation (4.8b) follows from (5.4) by cancellation. ||

Before we prove Theorem 4.7, we make two explanatory remarks.

Remark 5.3. Assumptions 4(i) and 4(ii) of Theorem 4.7 together imply that for $I \in \mathcal{I}$,

$$\lim_{n\to\infty} \sup_{0\le t<\alpha(F)} |F_n(\{t\}, I) - F(\{t\}, I)| = 0.$$

Remark 5.4. Remark 5.3 implies that for $I \in \mathcal{I}$,

$$D(I) \subset \bigcup_{n=1}^{\infty} \bigcap_{k\ge n} D(I, k).$$

The proof of Theorem 4.7 is based upon the following three lemmas.

Lemma 5.5. Under assumptions 4(i)-4(iv) of Theorem 4.7,

$$\alpha \equiv \varliminf_{n\to\infty} \alpha(F_n) \ge \alpha(F). \tag{5.5}$$

Proof. It suffices to consider the case $\alpha < \infty$. Let $\{m\}$ be an arbitrary infinite subsequence of $\{1, 2, \ldots\}$ such that $\lim_{m\to\infty} \alpha(F_m) = \alpha$. By definition, $\overline{F}_m[\alpha(F_m)] = 0$. By 4(iv), $\lim_{m\to\infty} \overline{F}[\alpha(F_m)] = 0$. Hence $\overline{F}(\alpha) = 0$. Relation (5.5) follows from the definition of $\alpha(F)$. ||

In Lemmas 5.6 and 5.7 below, let $\{m\}$ denote an arbitrary infinite subsequence of $\{1, 2, \ldots\}$, let $I \in \mathcal{I}$, and let $0 \le t \le \alpha(F)$.

Lemma 5.6. Let

$$B_{1,m}^{I}(t) = \sum_{a\in D(I)} \chi_{D(I,m)}(a)\, \chi_{[0,t)\,\cap\,[0,\alpha(F_m))}\, \ell n[\overline{F}_m(a)/\overline{F}_m(a^-)],$$

and

$$B_{1}^{I}(t) = \sum_{a\in D(I)} \chi_{D(I)}(a)\, \chi_{[0,t)\,\cap\,[0,\alpha(F))}\, \ell n[\overline{F}(a)/\overline{F}(a^-)].$$

Then

$$\overline{\lim_{m\to\infty}}\ B^I_{1,m}(t) \le B^I_1(t). \tag{5.6}$$

Proof. By Fatou's Lemma,

$$\overline{\lim_{m\to\infty}}\ B^I_{1,m}(t) \le \sum_{a\in D(I)} \overline{\lim_{m\to\infty}}\ \chi_{D(I,m)\,\cap\,[0,\,t)\,\cap\,[0,\alpha(F_m))}(a)\ \ln[\overline{F}_m(a)/\overline{F}_m(a^-)].$$

By Remark 5.3, $\lim_{m\to\infty} \chi_{D(I,m)}(a) = 1$ for $a \in D(I)$. By Lemma 5.5, $\lim_{m\to\infty} \chi_{[0,t)\,\cap\,[0,\alpha(F_m))}(a)$ $= \chi_{[0,t)}(a)$. Assumption 4(iv) implies that

$$\lim_{m\to\infty} \ln[\overline{F}_m(a)/\overline{F}_m(a^-)] = \ln[\overline{F}(a)/\overline{F}(a^-)].$$

Relation (5.6) follows. ||

Lemma 5.7. Let

$$B^I_{2,m}(t) = \sum_{a\in D(I,m)} \chi_{[0,t)\,\cap\,[0,\alpha(F_m))\,\cap\,C(I)}(a)\ \ln[\overline{F}_m(a)/\overline{F}_m(a^-)]$$

$$-\int_{[0,t)} [\chi_{((I,m)\,\cap\,[0,\alpha(F_m))}/\overline{F}]\ dF_m(\cdot,\ I)$$

and

$$B_2{}^I(t) = \sum_{a\in D(I)} \chi_{[0,t)\,\cap\,[0,\alpha(F))\,\cap\,C(I)}(a)\ \ln[\overline{F}(a)/\overline{F}(a^-)]$$

$$-\int_{[0,t)} [\chi_{C(I)\,\cap\,[0,\alpha(F))}/\overline{F}]\ dF(\cdot,\ I).$$

Then

$$\overline{\lim_{m\to\infty}}\ B^I_{2,m}(t) \le B_2{}^I(t). \tag{5.7}$$

<u>Proof</u>. Let $\varepsilon > 0$. By 4(i),

$$\sum_{a\in D(I,m)} \chi_{[0,t) \cap [0,\alpha(F_m)) \cap C(I)} \ln[\overline{F}_m(a)/\overline{F}_m(a^-)]$$

$$= -\int_{[0,t)} \chi_{D(I,m) \cap C(I) \cap [0,\alpha(F_m))}(u) \ln[1 + F_m(\{u\}, I)/\overline{F}_m(u)]^{1/F_m(\{u\}, I)} d\, F_m(u, I)$$

$$\leq \varepsilon -\int_{[0,t)} [\chi_{D(I,m) \cap C(I) \cap [0,\alpha(F_m))}/\overline{F}] \, d\, F_m(\cdot, I)$$

for m sufficiently large by 4(iv). Again by 4(iv),

$$-\int_{[0,t)} [\chi_{C(I,m) \cap [0,\alpha(F_m))}/\overline{F}_m] \, d\, F_m(\cdot, I)$$

$$\leq \varepsilon \; -\int_{[0,t)} [\chi_{C(I) \cap C(I,m) \cap [0,\alpha(F_m))}/\overline{F}] \, d\, F_m(\cdot, I).$$

Hence

$$B^I_{2,m}(t) \leq 2\,\varepsilon \; -\int_{[0,t-\delta]} [\chi_{C(I) \cap [0,\alpha(F_m))}/\overline{F}] \, d\, F_m(\cdot, I)$$

for $0 < \delta < t$. It follows from Lemma 5.5 that

$$\lim_{m\to\infty} \sup_{0\leq u\leq t-\delta} |\chi_{[0,t-\delta] \cap [0,\alpha(F_m))}(u) - \chi_{[0,t-\delta]}(u)| = 0.$$

Therefore, for m sufficiently large,

$$B^I_{2,m}(t) \leq 2\,\varepsilon + \varepsilon/\overline{F}(t) \; -\int_{[0,t-\delta]} [\chi_{C(I)}/\overline{F}] \, d\, F_m(\cdot, I).$$

Using the Helly-Bray Theorem, we conclude from 4(ii) that

$$\overline{\lim_{m\to\infty}} \, B^I_{2,m}(t) \leq \; -\int_{[0,t-\delta]} [\chi_{C(I)}/\overline{F}] \, d\, F(\cdot, I)$$

for every δ such that $0 < \delta < t$. Relation (5.7) is established by letting $\delta \downarrow 0$. ||

<u>Proof of Theorem 4.7</u>. Let $I \in \mathcal{I}$, $0 \le t \le \alpha(F)$, and $\{m\}$ be an arbitrary infinite subsequence of $\{1, 2, \ldots\}$. By (2.2),

$$\overline{G}_{I,m}(t^+) = \exp[B^I_{1,m}(t) + B^I_{2,m}(t)].$$

By Lemmas 5.6 and 5.7,

$$\overline{\lim_{m\to\infty}}\, \overline{G}_{I,m}(t^+) \le \exp[B^I_1(t) + B^I_2(t)] = \overline{G}_I(t^+).$$

By Theorem 2.1, we may conclude that

$$P(\tau_m \ge t) = \prod_{I\in\mathcal{I}} \overline{G}_{I,m}(t^+).$$

Thus,

$$\overline{F}(t^+) = \lim_{m\to\infty} \prod_{I\in\mathcal{I}} \overline{G}_{I,m}(t^+) \le \prod_{I\in\mathcal{I}} \overline{\lim_{m\to\infty}}\, \overline{G}_{I,m}(t^+) \le \prod_{I\in\mathcal{I}} \overline{G}_I(t^+) = \overline{F}(t^+).$$

Hence, $\overline{\lim_{m\to\infty}}\, \overline{G}_{I,m}(t^+) = \overline{G}_I(t)$. ||

REFERENCES

[1] Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. Scand. J. of Statist. 3, 15-27.

[2] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2, 437-453.

[3] Desu, M. M. and Narula, S. C. (1977). Reliability estimation under competing causes of failure. The Theory and Applications of Reliability I, ed. by I. Shimi and C. P. Tsokos, Academic Press: New York.

[4] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

[5] Langberg, N., Proschan, F. and Quinzi, A. J. (1978). Converting dependent models into independent ones, preserving essential features. Ann. Prob. 6, in press.

[6] Meier, P. (1975). Estimation of a distribution function from incomplete observations. Perspectives in Probability and Statistics, Applied Prob. Trust, Sheffield.

[7] Miller, D. R. (1977). A note on independence of multivariate lifetimes in competing risks models. Ann. Statist. 5, 576-579.

[8] Peterson, A. V., Jr. (1975). Nonparametric estimation in the competing risks problem. Stanford University Technical Report No. 13.

[9] Peterson, A. V., Jr. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. J. Amer. Statist. Assoc. 72, 854-858.

[10] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. USA 72, 20-22.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER 2. GOVT ACCESSION NO. FSU No. M438	3. RECIPIENT'S CATALOG NUMBER
4. TITLE Estimating Dependent Life Lengths, with Applications to the Theory of Competing Risks	5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M438
7. AUTHOR(s) N. Langberg, F. Proschan, A. J. Quinzi	8. CONTRACT OR GRANT NUMBER(s) AFOSR 74-2581D
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS The U.S. Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332	12. REPORT DATE November, 1977 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Competing risks, life distribution, simultaneous failure, reliability, Kaplan-Meier estimator, consistent estimator, survival probability, dependent life lengths, censored data, dependent components, estimating risks.

20. ABSTRACT

In the classical theory of competing risks (as well as in many reliability models and incomplete data problems) it is assumed that (a) the risks (i.e., the random variables of interest) are independent and that (b) death does not result from simultaneous causes. Employing our probabilistic solution to a related problem in probability modelling, we obtain strongly consistent estimators for the unobservable marginal distributions of interest. These estimators are analogous to those of Kaplan and Meier [J. Amer. Statist. Assoc. (1958) 63] and are appropriate when the assumptions of independence and no simultaneous causes of death [(a) and (b), above] fail to hold. We show how our methods can be used to unify and simplify the nonparametric approach toward estimation in the competing risks model. As a consequence we obtain an elementary proof of the strong consistency of the Kaplan-Meier estimator. Our results extend and simplify the work of Peterson [J. Amer. Statist. Assoc. (1977) 72] and Desu and Narula [The Theory and Applications of Reliability, I, ed. by I. Shimi and C. P. Tsokos (1977)].